

**IMAGE DEFORMATION METHOD BASED ON VOICE SIGNAL**

Patent Number: JP4359299  
Publication date: 1992-12-11  
Inventor(s): KAWATE FUMITAKA  
Applicant(s): SONY CORP  
Requested Patent: ☐ JP4359299  
Application Number: JP19910134954 19910606  
Priority Number(s):  
IPC Classification: G10L3/00; G09B5/02; G10L9/04  
EC Classification:  
Equivalents: JP3070136B2

**Best Available Copy****Abstract**

**PURPOSE:** To make a smooth transition of a face image deformation, which corresponds to the image while a certain vowel is pronounced, to another face image which corresponds to the image while a different vowel is pronounced.

**CONSTITUTION:** From an input voice signal, the control frequency of a formant frequency, which shows a peak of the spectrum envelope of the input voice signal, is computed, the formant frequency is linear and nonlinear transformed, at least two parameters of the lower jaw data and the lip data are obtained, from these two parameters at least two functions, i.e., a first function which corresponds to a first vowel group and a second function which corresponds to a second vowel group are obtained as the lower jaw opening and the lip opening functions and the shape of a face being displayed is deformed based upon the amount of change of the lower jaw and the lip obtained from these functions.

Data supplied from the esp@cenet database - I2

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平4-359299

(43) 公開日 平成4年(1992)12月11日

(51) Int.Cl. <sup>5</sup>	識別記号	序内整理番号	F I	技術表示箇所
G 1 0 L 3/00	S	8946-5H		
G 0 9 B 5/02		8603-2C		
G 1 0 L 9/04	Z	8946-5H		
// A 6 3 F 9/22	B	8804-2C		

審査請求 未請求 請求項の数1(全 7 頁)

(21) 出願番号 特願平3-134954

(22) 出願日 平成3年(1991)6月6日

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 川手 史隆

東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

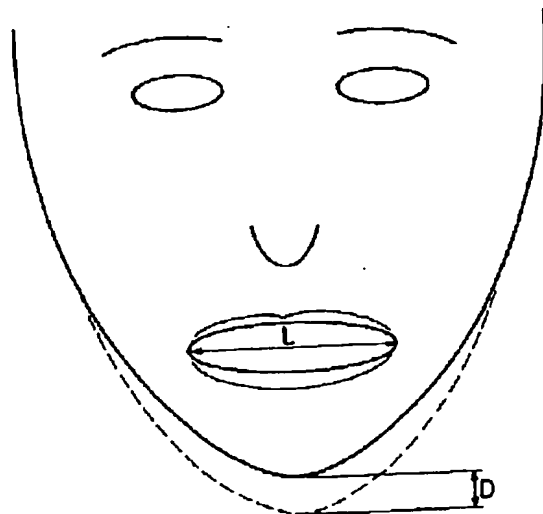
(74) 代理人 弁理士 松隈 秀盛

(54) 【発明の名称】 音声信号に基づく画像の変形方法

(57) 【要約】

【目的】 ある母音を発声している画像から別の母音を発声している画像へと言うような入力音声に対応した顔画像の変化が、スムーズになるようにする。

【構成】 入力音声信号から、当該入力音声信号のスペクトルエンベロープのピークを示すホルマント周波数の中心周波数を求め、このホルマント周波数を線型変換及び非線型変換することにより、下顎データ及び口唇データの少なくとも2つのパラメータを求め、これら2つのパラメータから第1の母音群に対応する第1の関数と、第2の母音群に対応する第2の関数との少なくとも2つの関数を、それぞれ下顎開大度及び口唇開大度における関数とし、これらの関数により求めた下顎及び口唇の変化量に基づき、画像表示した顔の形状を変形させるようにした。



顔 画 像

## 【特許請求の範囲】

【請求項1】 入力音声信号から、当該入力音声信号のスペクトルエンベロープのピークを示すホルマント周波数の中心周波数を求め、このホルマント周波数を線型変換及び非線型変換することにより、下顎データ及び口唇データの少なくとも2つのパラメータを求め、これら2つのパラメータから第1の母音群に対応する第1の関数と、第2の母音群に対応する第2の関数との少なくとも2つの関数を、それぞれ下顎開度及び口唇開度における関数とし、これらの関数により求めた下顎及び口唇の変化量に基づき、画像表示した顔の形状を変形させることを特徴とする音声信号に基づく画像の変形方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、音声信号に基づいて画像の顔の顎と口唇の形状を制御する音声信号に基づく画像の変形方法に関する。

## 【0002】

【従来の技術】 従来の例えばいわゆるアニメーションにおいて、そのアニメーション中の人物が会話などを行う際の口唇、顎などの顔の動きは、当該アニメーション画像の作成者が、このときの会話に合わせた口唇などの動きを例えば従来の経験に照らし合わせて推測することで決めるようにしている。

【0003】 ところで、近年、アニメーションなどにおいて、会話に合わせて、よりリアルに口唇、顎などを動かすことができるようにすることが求められている。

【0004】 このため、音声のホルマント成分を利用して、口唇などの動きを制御させることが考えられている。即ち、音声の特徴量の一つにホルマントがあり、母音の第1・第2ホルマントと発声時の舌面形状の曲率・下顎開度が対応付けられることが既に知られている。この方法は、この第1・第2ホルマントと顎・口唇の開度の対応付けを利用してアニメーションの制御を行うもので、音韻の認識を行なうことなく動画像の制御を行なうものである。即ち、図1に示すように、アニメーション

\* ヨンの顔画像の口唇の開度Lと下顎の開度Dとを制御するものである。

【0005】 ここで、ホルマントと顎・口唇の開度は線形の変換では完全には対応付けられないこと、顎・口唇の開度に加え、「イ」らしさのパラメータが必要ながことがわかった。

【0006】 ホルマント周波数は音声の基本的な音響パラメータの一つである。音声信号のスペクトル包絡を調べると、いくつかのピークがあることがわかる。このピークとなる周波数がホルマント周波数であり、周波数の低い方から順に第1ホルマント、第2ホルマントと呼ぶ。即ち、図11は母音「ア」の波形のパワースペクトルを示す図で、この図11に示すように、周波数の低い方から順に第1ホルマントF1、第2ホルマントF2・・・とピークがあるのが判る。一般にホルマントは声道の音響的インパルス応答の減衰正弦波成分と定義される。平均的声道に対しては一般に3kHz以内に3ないし4個のホルマントがある。なお、母音「イ」及び「ウ」の波形のパワースペクトルを、図12及び図13に示す。

【0007】 また、音声「ア・イ・ウ・エ・オ」の第1・第2ホルマント周波数の遷移の様子を図14に示す。このように第1・第2ホルマントを調べることで、音声信号がどの母音のものかをほぼ推定できる。

【0008】 そして、第1・第2ホルマント平面（以下、F1-F2平面）上と調音平面（舌面形状の曲率と下顎開度）上とでは、各母音の相対的な位置関係が一致する。

【0009】 そこで、まず最初に、この結果を拡張して口唇の開度と舌の形状とに強い相関があるものと仮定し、F1-F2平面から顎-口唇平面への変換を試みると、次式にしたがって、図15のF1-F2平面の「ア」、「イ」、「ウ」の3母音による三角形は、顎-口唇平面の「ア」、「イ」、「ウ」による三角形に線形変換することができる。

## 【0010】

## 【数1】

$$\begin{bmatrix} M_{lip} \\ M_{jaw} \end{bmatrix} - \begin{bmatrix} M_{lip-o} \\ M_{jaw-o} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \left( \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} - \begin{bmatrix} f_{1o} \\ f_{2o} \end{bmatrix} \right)$$

【0011】 ここで、 $M_{lip}$  は口唇の開度、 $M_{jaw}$  は顎の開度、 $f_1$ 、 $f_2$  は第1・第2ホルマントである。また、 $M_{lip-o}$ 、 $M_{jaw-o}$ 、 $f_{1o}$ 、 $f_{2o}$  はそれぞれ、「ウ」に対する口唇、顎の開度、第1・第2ホルマントであり、 $a$ 、 $b$ 、 $c$ 、 $d$  は定数である。

【0012】 この【数1】式を用いた変換により、各母音の顎-口唇平面上の位置は図16のようになる。この図16において、両パラメータの値が0のときは発声なしの状態であり、正の時は開いた状態、負の時はつぼめた状態を表す。

【0013】 この図16に示されるように線形に変換し

た場合には、「エ」は適当と思われる位置となるが、「オ」については、口唇を「ウ」よりもつぼめることになっていて不適当であり、これについての補正が必要である。しかし、各母音のホルマントの値がある幅を持って分布すること、したがって、その線形変換の結果も同様に分布することから、顎-口唇平面上でこの補正を行なうためにはその音韻を判別することが必要となり、あまり好ましくない。また、このパラメータがそのまま画像発生装置に利用できるのではなく、パラメータの変換が必要なこと、顔動画像の動きのスムージング処理が必要なことなどから、この補正は画像発生装置への入力デ

ータの算出に含めている。

【0014】

【発明が解決しようとする課題】ところで、公知の画像発生装置にて顔動画像の制御を行なう場合、口の動きに関するパラメータとしては顎を開く度合い、口唇をつばめる度合い、歯を見せる（上唇を上引っ張る）度合い（以下、順にM-JAW、M-LIP-P、M-LIP-I）の3つが定義されている。大雑把に言って、上述した〔数1〕式を用いて求めた顎と口唇の開大度はM-JAWとM-LIP-Pにそれぞれが対応する。

【0015】まず、ホルマントの線形変換により求めたデータ（以下、顎・口唇データ）をそのまま画像発生装置に入力して制御を行なったところ、「イ」や「エ」のような口の形を表現できず、また、「ア」の入力に対して「オ」と発声しているような画像となった。

【0016】次に、口唇データに対して適当なしきい値を設定し、そのしきい値と入力データを比較して入力「イ」であるか否かを判断してM-LIP-Iを生成するようにしたところ、「イ」は表現できるが、「ア」や「エ」は表現できなかった。さらに「ア」や「エ」まで判断することも試みたが、困難であり、また、合成された動画画像も動きがスムーズなものとはならなかった。

【0017】以上のように顎・口唇データをそのまま与えたのでは、「ウ」や「オ」のような丸い口の形は表現できるが、「イ」や「エ」のような四角い口の形を表現することができないため、顎・口唇データの2つのパラメータから、M-JAW、M-LIP-P、M-LIP-Iの3つのパラメータを導出しなければならないことがわかった。また、パラメータの決定に際して、ある母音から別の母音へ、無音から発声へなどの中間状態での動作を自然なものにするために、各パラメータは連続なデータとなるように求めなければならない。さらに、発声開始・終了時付近では顎・口唇データが必ずしも適当なものとはならない。場合によっては、同じ母音に対して正・負の間を大きく振動することもある。これに対するスムージングの処理も必要となっていた。

【0018】本発明の目的は、ある母音を発声している顔画像から別の母音を発声している顔画像へと言うような変化がスムーズになるように、中間状態の画像が良好に作成されるようにすることにある。

【0019】

【課題を解決するための手段】本発明は、入力音声信号から、当該入力音声信号のスペクトルエンベロープのピークを示すホルマント周波数の中心周波数を求め、このホルマント周波数を線型変換及び非線型変換することにより、下顎データ及び口唇データの少なくとも2つのパラメータを求め、これら2つのパラメータから第1の母音群に対応する第1の関数と、第2の母音群に対応する第2の関数との少なくとも2つの関数を、それぞれ下顎開大度及び口唇開大度における関数とし、これらの関数

により求めた下顎及び口唇の変化量に基づき、画像表示した顔の形状を変形させるようにしたものである。

【0020】

【作用】このようにしたことで、実際の音声に基づいた入力音声信号のホルマント周波数に、簡単な線型変換及び非線型変換の演算を施して、下顎開大度及び口唇開大度が求まると共に、下顎及び口唇の変化量が求まり、画像表示した顔の形状を入力音声に合わせてリアルタイムでスムーズに変化させることができるようになる。

10 【0021】

【実施例】以下、本発明の一実施例を、添付図面を参照して説明する。

【0022】本例においては、アニメーションの顔画像の口唇の開大度Lと下顎の開大度D（図1参照）とをスムーズに制御するようにしたもので、この下顎及び口唇の画像データを得るために、まず顎データM-JAW、口唇データM-LIP-P、口唇データM-LIP-Iの各平面を仮定する。口の形が最もそれらしく見えるようなところに各母音の位置を決める。

20 【0023】次に、各平面上で全ての母音の位置を通過する連続な曲線（関数）を定義する。この曲線は1次関数とsigmoid関数の組合せにより表現する。ここで、1つの曲線では全ての母音を結ぶのは難しいので、1つの平面に2つの曲線を定義し、2つの曲線により全ての母音を含むものとする。ただし、実際には演算の高速化と簡略化のためにsigmoid関数を直線近似したものを利用している。この曲線を図2、図3、図4に示す。この図2、図3、図4において、M-JAWとM-LIP-Iは値が大きいほど顎を開く度合い、歯を見せる度合いが大きい。M-LIP-Pについては値が正に大のときに口をつばめる度合いが大きく、負に大のときは口を横に開く度合いが大きいことを示す。

30 【0024】そして、M-JAWに対して口唇データ、M-LIP-Iに対して顎データというように、入力データが各平面上の2つの曲線のいずれに近いかを特定できるパラメータを基準として、それにより2つの曲線の重み付けを施した平均をとり、そのフレームの入力データに対する高速画像変換装置のパラメータとする。また、M-LIP-Pに対してはM-JAWの大きさに応じてM-LIP-Pの最大値を決定している（図5）。これは口唇をつばめた状態で同時に顎を大きく開くことはできないことを意味する。これにより、F1-F2平面から顎-口唇平面への線形変換では不適当な位置に写像されたデータの補正を行なっている。以上の処理により、M-JAW、M-LIP-P、M-LIP-Iの3つのパラメータは顎・口唇データにより一意的に求まり、かつ、顎・口唇データに対して連続な値を持つ。

40 【0025】さらに、現在のフレームを中心に前後2フレームずつ、合計5フレームのパラメータを求める。この5フレームでパラメータの値が大きく異なるものがあ

るとき（例えば、4フレームが正の値で、残りの1フレームが負の値のとき）にはそのフレームは無効とする。そして、現在のフレームに2、それ以外に1の重み付けをして有効なデータについて平均値を求め、これを画像発生装置に供給するデータとする。

【0026】なお、F1-F2平面から顎-口唇平面へ\*

母音	第1ホルマント [kHz]	第2ホルマント [kHz]	顎の開大度	口唇の開大度
ア	0.867	1.379	0.953	-0.078
イ	0.289	2.246	0.102	0.031
ウ	0.314	1.254	0.133	-1.0

【0028】このようにして、母音の離散発声と連続発声（ともに「ア」、「イ」、「ウ」、「エ」、「オ」の順に発声）のそれぞれについてデータを求め、同一の変換を行なって画像発生装置に供給した。その結果、離散発声・連続発声とも完全に動画の制御を行なうことができた。また無音から発声へ、発声から無音への変化もスムーズなものとなった。本例での顔動画像の作成例を図6、図7、図8、図9、図10に示す。この図6～図10は、「ア」、「イ」、「ウ」、「エ」、「オ」の発声時の顔動画像を示す。

【0029】本実施例では3つのパラメータだけで口の動きの制御を行っており、動きは左右対称なものとなっている。また、画像の変形は画像を構成する曲面の曲げ延ばしや伸縮によりなされる。

【0030】なお、動きのパラメータの求め方であるが、上述したように口の動きには左右非対称な成分も含まれている。上述実施例のように母音のみの表現においては左右対称な動きでも良いであろうが、顔全体の動きを制御して表情を表現しようとするならば、より高度な制御が必要になる。

【0031】また、本例のデータに基づいて画像発生装置で作成した画像は、曲面の伸縮・曲げ延ばしによる画像の変形であり、口の動きが口の周辺にも影響し、自然な動きに近いアニメーションとなる。

【0032】また、従来の方法では、音声入力による顔動画像の制御を行なう場合に音韻の認識を必要とするのに対し、本実施例では音韻の認識を行なうことなく画像の制御を行っており、音声の入力から顔動画像の出力までを考えた場合に、より高速に処理できる。

【0033】

【発明の効果】本発明によると、音声信号のホルマント周波数を与えると比較的簡単なアルゴリズムによって、画像の口の動きをスムーズに制御することが可能となる。従って、音声による動画像の制御の利用分野として、

・発話障害者の発声矯正への利用

\*の変換は「ア」、「イ」、「ウ」について次に示す【表1】のように対応付けられるように係数を求めて、変換を行なった。

【0027】

【表1】

・アニメーション作成支援

・言語学習用CAI

・情報検索システムのMMI

などに応用することができる。

【図面の簡単な説明】

【図1】本発明の一実施例により作成されるデータを示す説明図である。

20 【図2】一実施例による顎データからパラメータへの変換状態を示す特性図である。

【図3】一実施例による口唇データからパラメータへの変換状態を示す特性図である。

【図4】一実施例による口唇データからパラメータへの変換状態を示す特性図である。

【図5】一実施例によるパラメータの決定状態を示す特性図である。

【図6】一実施例により作成された画像例を示す説明図である。

30 【図7】一実施例により作成された画像例を示す説明図である。

【図8】一実施例により作成された画像例を示す説明図である。

【図9】一実施例により作成された画像例を示す説明図である。

【図10】一実施例により作成された画像例を示す説明図である。

【図11】音声の波形の一例を示す波形図である。

【図12】音声の波形の一例を示す波形図である。

40 【図13】音声の波形の一例を示す波形図である。

【図14】ホルマント周波数の遷移状態を示す特性図である。

【図15】F1-F2平面を示す特性図である。

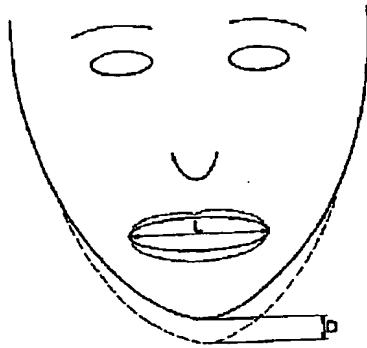
【図16】顎-口唇平面を示す特性図である。

【符号の説明】

D 下顎の開大度

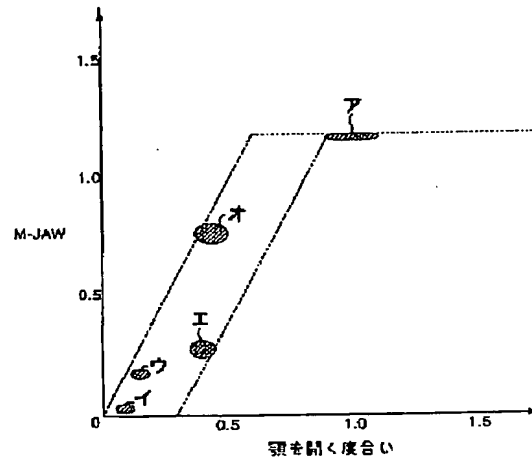
L 口唇の開大度

【図1】



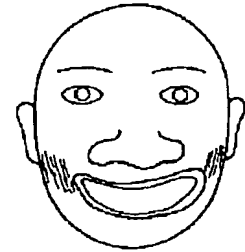
顔画像

【図2】



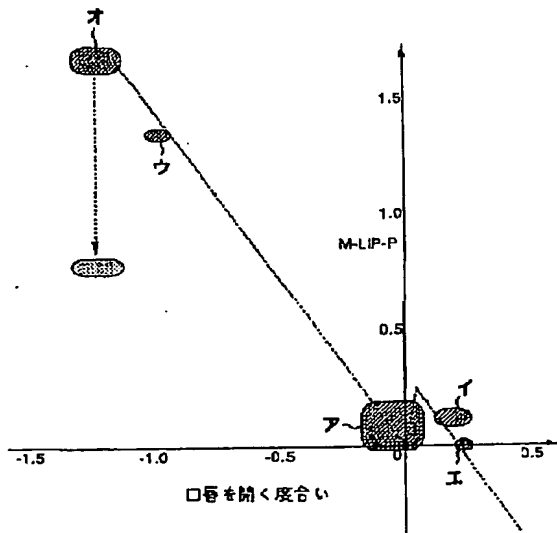
顎データからパラメータM-JAWへの変換

【図9】



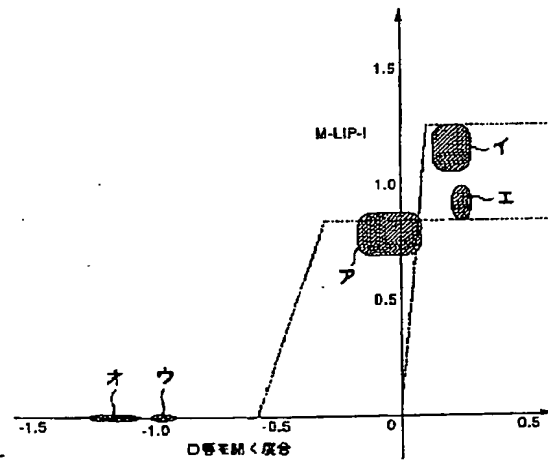
エの入力に対する顔画像

【図3】



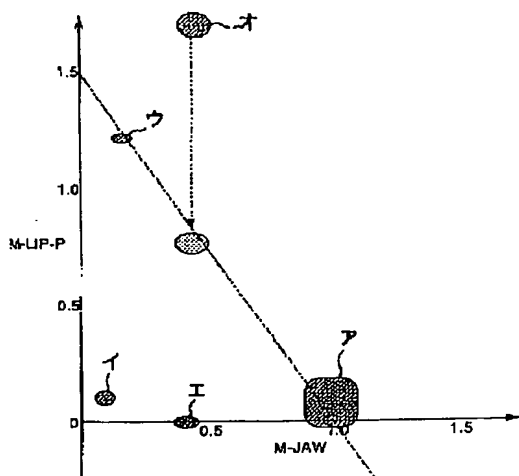
口唇データからパラメータM-LIP-Pへの変換

【図4】



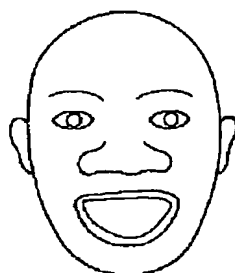
口唇データからパラメータM-LIP-Iへの変換

【図5】



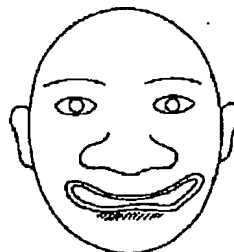
M-LIP-Pの最大値の決定

【図6】



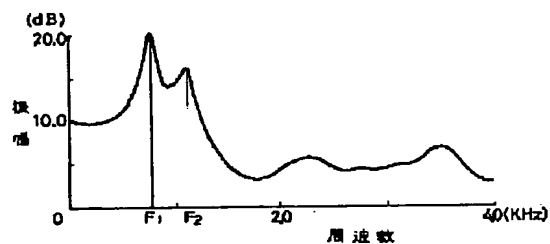
アの入力に対する顔画像

【図7】



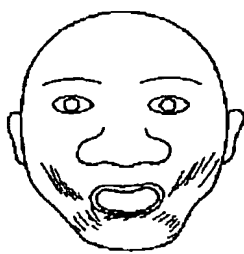
イの入力に対する顔画像

【図11】



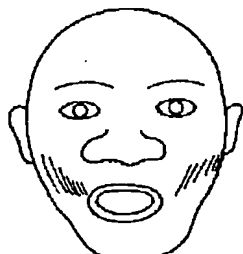
母音「ア」の波形

【図8】



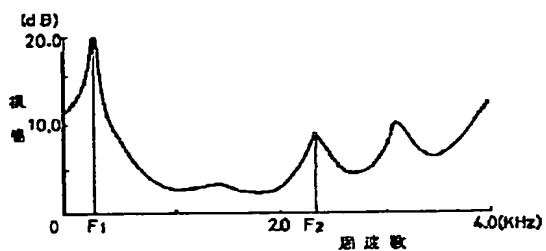
ウの入力に対する顔画像

【図10】



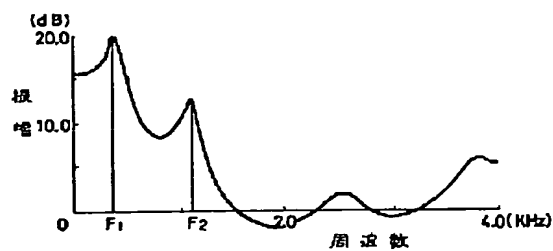
オの入力に対する顔画像

【図12】



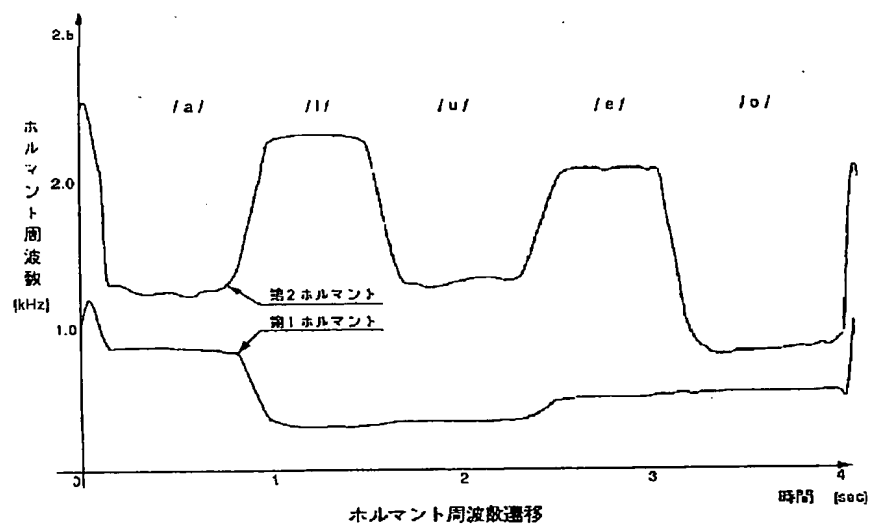
母音「イ」の波形

【図13】

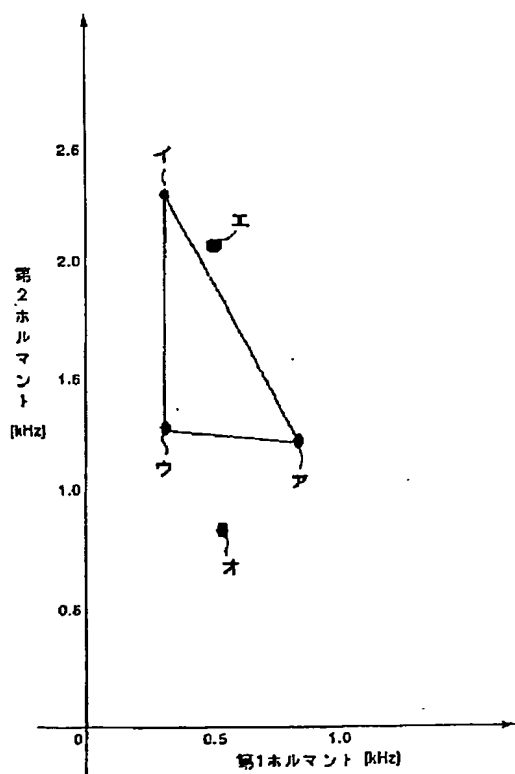


母音「ウ」の波形

【図14】

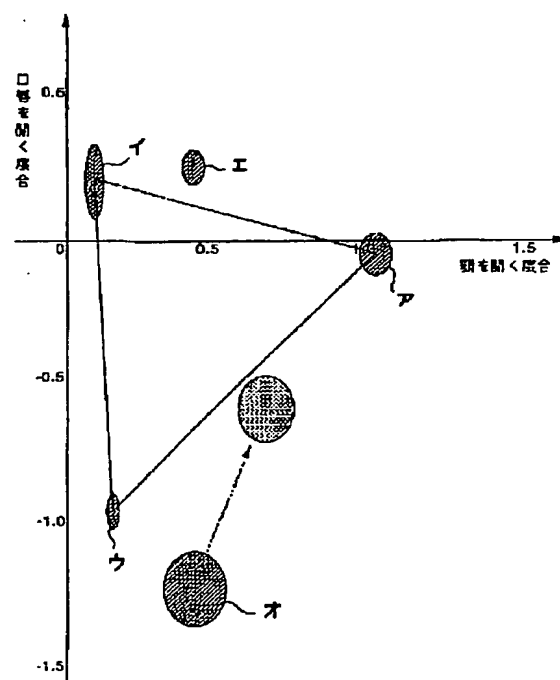


【図15】



F1-F2 平面

【図16】



顎-口唇平面



【公報種別】特許法第17条の2の規定による補正の掲載  
 【部門区分】第6部門第2区分  
 【発行日】平成10年(1998)12月8日

【公開番号】特開平4-359299  
 【公開日】平成4年(1992)12月11日  
 【年通号数】公開特許公報4-3593  
 【出願番号】特願平3-134954  
 【国際特許分類第6版】

G10L 3/00  
 G09B 5/02  
 G10L 9/04  
 // A63F 9/22

【F I】

G10L 3/00 S  
 G09B 5/02  
 G10L 9/04 Z  
 A63F 9/22 B

【手続補正書】  
 【提出日】平成10年6月8日  
 【手続補正1】  
 【補正対象書類名】明細書  
 【補正対象項目名】特許請求の範囲  
 【補正方法】変更  
 【補正内容】  
 【特許請求の範囲】

【請求項1】 入力音声信号から、当該入力音声信号のスペクトルエンベロープのピークを示すホルマント周波数の中心周波数を求め、このホルマント周波数を線型変換及び非線型変換することにより、下顎及び口唇の動きに関する少なくとも2つのパラメータを求め、  
 これら2つのパラメータから第1の母音群に対応する第1の関数と、第2の母音群に対応する第2の関数との少なくとも2つの関数を、それぞれ下顎開大度及び口唇開大度における関数とし、  
 これらの関数により求めた下顎及び口唇の変化量に基づき、画像表示した顔の形状を変形させることを特徴とす

る音声信号に基づく画像の変形方法。

【手続補正2】  
 【補正対象書類名】明細書  
 【補正対象項目名】0019  
 【補正方法】変更  
 【補正内容】  
 【0019】

【課題を解決するための手段】本発明は、入力音声信号から、当該入力音声信号のスペクトルエンベロープのピークを示すホルマント周波数の中心周波数を求め、このホルマント周波数を線型変換及び非線型変換することにより、下顎及び口唇の動きに関する少なくとも2つのパラメータを求め、これら2つのパラメータから第1の母音群に対応する第1の関数と、第2の母音群に対応する第2の関数との少なくとも2つの関数を、それぞれ下顎開大度及び口唇開大度における関数とし、これらの関数により求めた下顎及び口唇の変化量に基づき、画像表示した顔の形状を変形させるようにしたものである。

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**